CHAPTER 1

MODIFIED AND HYBRID CONJUGATE GRADIENT METHODS WITH THEIR CONVERGENCE ANALYSIS: A REVIEW

Abstract

Conjugate Gradient (CG) methods are widely used for solving unconstrained optimization problems. The paper reviews the development CG methods in recent times and their convergence properties.

1.1 Introduction

Conjugate Gradient (CG) methods are among the earliest known techniques for solving large scaled unconstrained optimization problems. The methods comprise a class of unconstrained optimization algorithms which are characterized by low memory requirements and strong, local, and global convergence properties respectively. These properties make them attractive to mathematicians and engineers for solving large scaled problems (Lu *et al.*, 2015). The work in Hestenes and Stiefel (1952) presented CG algorithm for solving symmetric, positive-definite linear systems. CG methods have applications in many fields of endeavours, such as control science, engineering, management science and operations research. We will consider the following optimization problem given in (1.1)

$$min\{f(x) : x \in \mathbb{R}^n\} \tag{1.1}$$

Where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, bounded from below. The iterative formula of a CG method is given by (1.2)

$x_0 \in \mathbb{R}^n$

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

Where α_k is a step length to be computed by a line search procedure and d_k is the search direction defined by (1.3)

$$d_0 = -g_0, \qquad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, ... \tag{1.3}$$

Where $g_k = \nabla f(x_k)$, is a column vector and β_k is a scalar called the CG update parameter. The choice of β_k is the major difference among some CG methods such as Fletcher-Reeve's(FR), the Hestenes-Stiefel (HS), the Polak-Ribiere-Polyak (PRP), the Conjugate Descent (CD), the Liu-Storey (LS), and the Dai-Yuan (DY), Dai and Yuan (2001a); Fletcher and Reeves (1964); Hestenes and Stiefel (1952); Hu and Storey (1991a); Polak and Ribiere (1969). The construction of algorithm by Fletcher-Reeves in 1964 is considered the first nonlinear CG method since it is on nonlinear optimization. Hestenes and Stiefel (1952) presented a CG algorithm for solving symmetric, positive-definite linear system and Fletcher and Reeves (1964) presented the CG method for solving nonlinear optimization problems. Let $||.||$ denote the Euclidean norm and define $y_k = g_{k+1} - g_k$, the parameter β_k of some CG methods are given in table 1.1

Table 1.1: The parameter β_k

S/N	β_k	**Method name**	**References**
1	$\frac{\|\|g_{k+1}\|\|^2}{\|\|g_k\|\|^2}$	Fletcher-Reeves(FR) method	Fletcher and Reeves (1964)
2	$-\frac{\|\|g_{k+1}\|\|^2}{d_k^T g_k}$	Conjugate descent(CD) method	Fletcher (1987)
3	$-\frac{\|\|g_{k+1}\|\|^2}{d_k^T y_k}$	Dai-Yuan(DY) method	Dai and Yuan (1999)
4	$\frac{g_{k+1}^T y_k}{\|\|g_k\|\|^2}$	Poyak-Rebiere-polak(PRP) method	Polyak (1969)
5	$-\frac{g_{k+1}^T y_k}{d_k^T g_k}$	Liu-Storey(LS) method	Hu and Storey (1991a,b)
6	$\frac{g_{k+1}^T y_k}{d_k^T g_k}$	Hestenes-Stiefel(HS) method	Hestenes and Stiefel (1952)
7	$(y_k - 2d_k \frac{\|\|y_k\|\|^2}{d_k^T y_k}) \frac{g_{k+1}}{d_k^T y_k}$	Hager and Zhang method	Hager and Zhang (2005)

Meanwhile if f is a convex quadratic function, the methods above are equivalent (Yuan *et al.*, 1999) and α_k is calculated by the exact line search, their behaviors for general objective functions may be far different. It is observed that the numerator of parameter β_k in table 1.1 is either $||g_{k+1}||^2$ or $g_{k+1}^T y_k$ and the denominator is either $||g_k||^2$ or $d_k^T y_k$ or $-d_k^T g_k$ except for the last parameter β_k. For non-quadratic cost functions, each choice for the update parameter leads to different performance. Some of today's best performing CG algorithms are hybrid methods which normally obtained by adjusting β_k as the iteration evolve, and a method based on the recent update of parameter β_k^N, with close connection to memory-less quasi-Newton methods.

1.2 Line Search

Let$f : \mathbb{R}^n \to \mathbb{R}^n$ be given and suppose that x_k is the current best estimate of a solution to (1.4)

$$min_{\alpha \geq 0} f(x_k + \alpha d_k). \tag{1.4}$$

A standard method for improving the estimate x_k is to choose a direction of search $d \in R^n$ and the compute a step length $\alpha^* \in R$ so that $x_k + \alpha^* d_k$ approximately optimizes f along the line $x + \alpha d | \alpha \in R$. The new estimate for the solution to (1.4) is then $x_{k+1} = x_k + \alpha^* d_k$. The process of choosing α^* is called a line search method.

In each CG iteration, the step size α_k is chosen to bring to an approximate minimum for the problem given in(1.4). Since $\alpha \geq 0$, the direction d_k should satisfy the descent condition in (1.5) for all $k \geq 0$. If there exist a constant $c \geq 0$ such that (1.6) holds, then the search directions satisfy the sufficient descent condition.

$$g_k^T d_k \leq 0, \tag{1.5}$$

$$g_k^T d_k \leq -c||g_k||^2 \tag{1.6}$$

Generally, for the convergence analysis and implementation of CG methods, the step size α_k is usually obtained by exact line search or inexact line searches. The inexact line searches which are classified as standard Wolfe condition (Wolfe conditions), the strong Wolfe conditions or the strong *Wolfe conditions(generalized Wolfe condition), which are as follows:

1. The Standard Wolfe line search is to find α_k such that

$$\begin{cases} f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \end{cases} \tag{1.7}$$

$$\begin{cases} d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k, \end{cases} \tag{1.8}$$

with $0 < \delta < \frac{1}{2}$ and $\delta < \sigma < 1$

2. The strong Wolfe line search is to find α_k such that

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \end{cases} \tag{1.9}$$

$$\begin{cases} |d_k^T g(x_k + \alpha_k d_k)| \leq -\sigma d_k^T g_k, \end{cases} \tag{1.10}$$

with $0 < \delta < \frac{1}{2}$ and $\delta < \sigma < 1$

3. The strong *Wolfe (generalized) line search is to find $alpha_k$ such that

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \end{cases} \tag{1.11}$$

$$\begin{cases} \sigma_1 d_k^T g_k \leq d_k^T g(x_k + \alpha_k d_k) \leq -\sigma_2 g_k^T d_k, \end{cases} \tag{1.12}$$

with $0 < \delta < \sigma_1 < 1$ and $\sigma_2 \geq 0$ and where d_k is a descent direction.

Observed that, for all the three Wolfe line searches, the differences are in (1.8), (1.10) and (1.12). The special case where $\sigma_1 = \sigma_2 = \sigma$ corresponds to the strong Wolfe conditions. Usually line search is terminated in a CG algorithm when the standard Wolfe conditions are satisfied. For some CG algorithms, however, stronger versions of the Wolfe conditions are needed to ensure convergence. Along the line an approximate Wolfe conditions was introduced as given in (1.13) and (1.14)

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \tag{1.13}$$

$$\sigma g_k^T d_k \leq g_{k+1}^T d_k \leq (2\sigma - 1) g_k^T d_k, \tag{1.14}$$

where $0 < \delta < \frac{1}{2}$ and $\delta < \sigma < 1$. The Wolfe condition in (1.14) is same as (1.7) if f is quadratic. In general, when $\pi(\alpha) = f(x_k + \alpha d_k)$ is replaced by a quadratic interpolant $q(.)$ that matches $\phi(\alpha)$ at $\alpha = 0$ and $\phi^{'}(\alpha)$ at $\alpha = 0$ and $\alpha = \alpha_k$, (1.7) is same as (1.14). Observed that with special choice for σ_2 in (1.12) differs from approximate Wolfe conditions. The standard, generalized, strong Wolfe conditions or exact line search are use to prove convergence of CG methods. The approximate Wolfe conditions are used in efficient, high accuracy implementations of CG algorithms for which there is no convergence theory, but the practical performance is most at times much better. As shown by (Hager and Zhang, 2005), the first Wolfe condition (1.7) limits the accuracy of a CG algorithm to the order of the square root of the machine precision, while with the approximation contained in (1.14),the accuracy is achieved on the order of the machine precision (Hager and Zhang, 2005). As explained further in (Hager and Zhang, 2006), faster convergence is achieved when using the approximate Wolfe conditions since a local minimizer of ϕ satisfied (1.14), while a point satisfying the standard or Wolfe conditions is obtained by computing a local minimizer of the approximating function ψ introduced in (Moré and Thuente, 1994) given in (1.15)

$$\psi(\alpha) = \phi(\alpha) - \phi(0) - \alpha\sigma\phi^{'}(0) \tag{1.15}$$

. When using the approximate line search, the function f along the search direction d_k is minimized rather than an approximation ψ to f. The approximate Wolfe search performed better in the computations but the global convergence of the algorithm cannot be guaranteed in theory. Meanwhile Dai and Kou (2013) proposed a modified line search called improved Wolfe line search. A positive sequence $\{\eta_k\}$ satisfying $\sum_{k\geq 1}\eta_k < +\infty$, given $\epsilon > 0$ we have (1.16)

$$f(x_k + \alpha_k d_k) \leq f(x_k) + min\left\{\epsilon|g_k^T d_k|, \delta\sigma_k g_k^T d_k + \eta_k\right\} \tag{1.16}$$

where δ and σ satisfying $0 < \delta < \sigma < 1$. Thus, called the line search satisfying (1.16) and (1.8) the improved Wolfe line search. It argued that (1.16) allows the step-sizes satisfying (1.7) and therefore is an extension of the standard Wolfe line search. If the trial point near x_k, in which case (1.17),we switch to requiring (1.18)

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \epsilon|f(x_k)|, \tag{1.17}$$

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta\sigma_k g_k^T d_k + \eta_k \tag{1.18}$$

instead of (1.8). The term η in (1.18) or (1.16) gives room for a slight increase in the function value and thereby avoiding the computational drawback of the standard line search (1.7). Also, the condition that the sequence $\{\eta_k\}$ be summable can guarantee global convergence of the algorithm to standard Wolfe line search.

In convergence analysis, either of the following assumptions on the objective function $f(x)$ are made

(i) The level set $\Omega = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded, where $x_0 \in \mathbb{R}^n$ is a given point.

(ii) In a neighbourhood M of Ω, f is continuously differentiable and its gradient g is

Lipschitz continuous, namely, there exists a constant $L > 0$ such that we have in (1.19)

$$||g(x) - g(y)|| \leq L||x - y||, \quad for \quad all \quad x, y \in M \tag{1.19}$$

The global convergence of the CG methods is proved by the following theorem usually referred to as Zoutendijk condition, the theorem was given by Zoutendijk (1970), Wolfe (1969, 1971).

Theorem 1 *Consider any CG method of the form $x_{k+1} = x_k + \alpha_k d_k$ where d_k satisfies $g_k^T d_k < 0$ for $k \in N^+$ and α_k satisfies the standard Wolfe line search. Then*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{||d_k||^2} < +\infty \tag{1.20}$$

The Global convergence proofs for CG methods are mostly based on the Zoutendijk condition combined with analysis showing that:
(a) the sufficient descent condition $g_k^T d_k \leq -c||g_k||^2$ holds and
(b) there exists a constant β such that $||d_k||^2 \leq \beta k$. (a), (b), and (1.18) yield

$$\liminf_{k \to \infty} ||g_k|| = 0 \tag{1.21}$$

A related to the Zoutendijk condition, found in Dai *et al.* (2000), is where the search directions are descent

Theorem 2 *Consider any CG method of the form $x_{k+1} = x_k + \alpha_k d_k$ where d_k satisfies $g_k^T d_k < 0$ for $k \in N^+$ and α_k satisfies the strong Wolfe conditions. If the Lipschitz assumption holds, then either*

$$\liminf_{k \to \infty} ||g_k|| = 0 \tag{1.22}$$

or

$$\sum_{k=1}^{\infty} \frac{||g_k||^4}{||d_k||^2} < \infty \tag{1.23}$$

The Wolfe line search and descent search direction conditions are independent of each other. Therefore, thereis need to satisfy some version of the Wolfe conditions, and make sure that the new search direction is a descent direction. The descent search direction condition holds automatically for the CG methods with the choice β_k^{DY} , when the line search satisfies the standard Wolfe conditions (Andrei, 2008b; Dai and Yuan, 1999; Lu *et al.*, 2015). And in the recent CG DESCENT (Babaie-Kafaki and Ghanbari, 2014; Dai and Kou, 2013; Dai and Yuan, 1999), sufficient descent holds for x_{k+1} with c = 7/8 if $d_k^T y_k \neq 0$.

1.3 The CG Methods with choices of β_k

Generally, Lipschitz Assumption is sufficient requirement of the global convergence of the CG iterative methods that have common numerator $||g_{k+1}||^2$ in theory such as the FR, DY and CD methods while PRP, LS, HS, among others requires both Lipschitz and Boundedness Assumption. The global convergence result given by Zoutendijk (1970) proved that FR method converges globally with exact line search. On the contrary Powell (1977) investigated that FR method with exact line search can cycle infinitely without making significant progress to the solution. The first global convergence analysis of the FR method with inexact line search was given by Al-Baali (1985). He also proved that FR method generates descent directions under the strong Wolfe conditions with $\sigma < \frac{1}{2}$ under strong Wolfe condition. With that, he proved that in (1.24)

$$\frac{1 - 2\sigma + \sigma^{k+1}}{1 - \sigma} \leq \frac{-g_k^T d_k}{||g_k||^2} \leq \frac{1 - \sigma^{k+1}}{1 - \sigma}, \tag{1.24}$$

for all $k \geq 0$. Dai and Yuan (1999) showed that the CD method always produces a descent direction if the strong Wolfe conditions are satisfied. The global convergence of CD method was proved under strong Wolfe line search with a strong restriction on the parameters. Notice that, with exact line search $\beta_k^{FR} = \beta_k^{CD}$. The major difference that exists between FR and CD is that for CD, sufficient descent (1.6) holds for strong Wolfe line search. The CD method does not need σ which was identified with FR. Furthermore, if a line search satisfies the generalized (strong *Wolfe) condition (1.12) with $\sigma_1 < 1$ and $\sigma_2 = 0$, then $0 \leq \beta_k^{CD} \leq \beta_k^{FR}$ follows and by Al-Baali (1985), global convergence is obtained. On the contrary, if $\sigma \geq 1$ or $\sigma > 0$ Dai and Yuan (1996) showed that $||d_k||^2$ increase exponentially through example given and the CD method converges to a point where the gradient does not vanish. Generally, DY method always generates descent direction with the standard Wolfe line search and it is globally convergent with the Lipschitz Assumption. This method first came in to being in Dai and Yuan (1999). Dai (2001) investigated the DY method further and established some remarkable properties, relating the descent directions of DY method to the sufficient descent condition:

Theorem 3 *Consider any CG method of the form $x_{k+1} = x_k + \alpha_k d_k$ with $d_{k+1} = -g_{k+1} + \beta_k d_k, \quad d_0 = -g_o$, where $\beta_k = \beta_k^{DY}$. If the DY method is implemented with any line search for which the search directions are descent directions, and if there exist constants γ_1 and γ_2 such that $\gamma_1 \leq ||g_k|| \leq \gamma_2 \quad for \quad all \quad k \geq 0$, then for any $p \in (0,1)$, there exists a constant $c > 0$ such that the sufficient descent given by (1.25)*

$$g_i^T d_i \leq -c||g_i||^2 \tag{1.25}$$

holds for all at least [pk] indices $i \in [0,k]$, where [j] denotes the largest integer $\leq j$.

Furthermore, Dai and Yuan established a convergence result applicable to any CG method where β_k can be expressed as in (1.26). Taking $\Phi_k = ||g_k||^2$ and $\Phi_{k+1} = ||g_{k+1}||$ yield β_k choice for FR method. From (1.3), we have (1.27).For exact line search $d_k^T g_{k+1} = 0$ and hence β_k^{DY} can be expressed as (1.28). The parameter β_k^{DY} of DY method correspond to (1.26) with $\Phi = g_k^T d_k$. The relation (1.26) play a key role in the convergence analysis of DY method.

$$\beta_k = \frac{\Phi_{k+1}}{\Phi_k}. \tag{1.26}$$

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \beta_k d_k^T g_{k+1} \tag{1.27}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{1.28}$$

In conclusion, all CG methods with $||g_{k+1}||^2$ in the numerator of β_k, that is; FR, CD and DY methods have strong global convergence properties (Al-Baali, 1985; Bai, 2001; Dai *et al.*, 2000; Dai, 2001; Guanghui *et al.*, 1995; Hu and Storey, 1991b). In practical computations, the nonlinear CG methods with $||g_{k+1}||^2$ in the numerator, though may have strong convergence but they may take infinitely small steps without approaching the solution. In that regards, methods such as PRP, HS and LS perform similarly theoretically and are preferred to the methods with $||g_{k+1}||^2$ in the numerator. The PRP, HS and LS are methods with $g_{k+1}^T y_k$ and they possess a built-in restart feature thereby addressing jamming problem of the earlier. The factor $y_k = g_{k+1} - g_k$ in the numerator of β_k tends to zero when the step $x_{k+1} - x_k$ becomes small thereby making

β_k small and the new search direction d_{k+1} becomes approximately steepest descent direction $-g_{k+1}$.

Polak and Ribiere (1969) proved the global convergence of the PRP method with exact line search using the assumption that f is strongly convex. As a consequence, Powell (1986) constructed a counter example which could not converge to a minimum using exact line search in the case of PRP and HS. Hence, the step size tends to zero, assumption required for convergence. The work of Powell (1986) motivated Gilbert and Nocedal (1992) to conduct an interesting analysis showing that PRP method converge globally if β_k^{PRP} is restricted to be non-negative and step size α_k is determined by a line-search step satisfying the sufcient descent condition $g_k^T d_k \leq -c||g_k||^2$ together with the standard Wolfe conditions. However, with the generalized line search condition (Grippo and Lucidi, 1997) proved that the PRP method is globally convergent for non-convex minimization. Also, Dai (1997) constructed an example showing that even when the objective function is strongly convex and $\sigma \in (0,1)$, the PRP method may still not converge with strong line search condition. From the work of Powell (1984), the modified update parameter for the PRP method was given in (1.29)

$$\beta_k^{PRP+} = max\{\beta_k^{PRP}, 0\}, \tag{1.29}$$

and its convergence was proved to curtail the convergence failure of PRP method with the Wolfe method (Gilbert and Nocedal, 1992). In order solve the problem of convergence failure of the method is to keep update parameter formula of PRP but modified the line search, where Grippo and Lucidi (1997) proposed a new Armijo type line search condition given by (1.30), where $j \geq 0$ is the smallest integer with the property in (1.31) and (1.32)

$$\alpha_k = max\{\lambda^j \frac{\tau|g_k^T d_k|}{||d_k||^2}\}, \tag{1.30}$$

$$f(x_{k+1}) \leq f(x_k) - \sigma\alpha_k^2||d_k||^2, \tag{1.31}$$

$$-\sigma_1||g_{k+1}||^2 \leq g_{k+1}^T d_{k+1} \leq -\sigma_2||g_{k+1}||^2 \tag{1.32}$$

where $0 < \sigma_2 < 1 < \sigma_1$, $0 < \lambda < 1$ and $\tau > 0$ are constants. The global convergence of PRP was achieved with the new line search condition. Dai *et al.* (2000) investigated global convergence with the line search take the step size $\alpha = \eta < \frac{1}{4L}$, where L is a Lipschitz constant for ∇f. In the same vain Sun and Zhang (2001) gave $\alpha_k = -\delta \frac{g_k^T d_k}{d_k^T Q_k d_k}$, where Q_k is some positive definite matrix with eigenvalue $v_{min} > 0, \delta \in (0, v_{min}|L)$, and L is a Lipschitz constant for ∇f In the case of HS method, the conjugacy condition $d_{k+1}^T y_k = 0$ holds, irrespective of the line search and

$\beta_k^{HS} = \beta_k^{PRP}$ with the exact line search.

Moreover, the two methods should have similarity in the convergence analysis. From the literature, the global convergence of the original HS and LS methods has not been proved under the mentioned line searches. Though, Shi and Shen (2007) proposed a new form of Armijo-type line search that guaranteed the global convergence of the LS method with the assumption of Lipschitz continuous partial derivatives. It needs to estimate the local Lipschitz constant of the derivative of objective functions in practical computation. The global convergence and linear convergence rate of the LS method with the new Armijo-type line search were analyzed under some mild conditions. While proving the global convergence of LS method using the new Armijo type line search, angle property was used in (1.33)

$$cos(-g_k, d_k) = -\frac{g_k^T d_k}{||g_k||.||d_k||} \geq \tau \tag{1.33}$$

where $1 \leq \tau > 0$ The Armijo line search was given as: Let $s > 0$ be a constant, $\rho \in (0, 1)$ and $\mu \in (0, 1)$. Choose α_k to be the largest α in $\{s, s\rho, s\rho^2, \cdots\}$ such that (1.34)

$$f_k - f(x_k + \alpha d_k) \geq -\alpha \mu g_k^T d_k \tag{1.34}$$

1.4 The Modified CG Methods

It has been shown through different researches that strong global convergence of FR, DY and CD methods are established. That is, the methods have strong global convergence but their computational performance is not so well in practice due to jamming phenomenon (Al-Baali, 1985; Dai, 2001; Dai and Yuan, 1999). On the other hand, HS, PRP and LS methods may not always converge but they often perform better, computational wise (Andrei, 2008b; Dai and Yuan, 1996). In order to take advantage of the features of the two categorizes of the CG methods which are strong global convergence and promising numerical computations, researchers proposed modification of some methods based on the β_k choice, hybridization of two families of the methods and descent direction adjustment. Liu and Feng (2011) suggested a modification of LS method given in (1.35)

$$\beta_k^{MLS} = \frac{g_k^T(g_k - t_k g_{k-1})}{u|d_{k-1}^T g_k| - d_{k-1}^T g_{k-1}}, \tag{1.35}$$

where $t_k = \frac{||g_k||}{||g_{k-1}||}, u > 0$. The numerical experiment showed that the modified method is efficient under the Grippo-Lucidi line search. Another Modified method proposed by Zhang *et al.* (2012) was based on PRP method in (1.36), where $\mu_1 \in (0, +\infty), \mu_2 \in (2\mu_1, +\infty), \mu_3 \in [\epsilon, +\infty)$ and epsilon is positive constant. This algorithm was implemented without line search and the global convergence was established with the strong Wolfe condition in (1.9) and (1.10) where $\delta \in (0, 1), \sigma \in (\delta, 1)$.

$$\beta_k^{MPRP} = \frac{\mu_1(||g_k||^2 - \frac{|g_k^T g_{k-1}|}{||g_{k+1}||^2} g_k^T g_{k-1})}{\mu_2 |g_k^T d_{k-1}| + \mu_3 ||g_{k-1}||^2} \tag{1.36}$$

Meanwhile, MPRP method converges globally for the general non-convex unconstrained optimization but the original PRP method with exact line may not. In the same manner the works of Dai and Wen (2012); Yuhong (2002) motivated Jiang and Jian (2013) to proposed another modified CG method called MDY whose intention was to improve the numerical performance that has a better property of the DY method and retain its good properties and structure. The modification was to the denominator of the original DY method in order to have the sufficient descent condition and also other properties for efficient conjugate gradient. The modified β_k given by (1.37), where $\mu > 1$. Also, by extension to FR method in the same paper, formula for β_k was proposed in (1.38)

$$\beta_k^{MDY} = \frac{||g_k||^2}{max\{d_k^T y_k, \mu |g_{k+1}^T d_k|\}} \tag{1.37}$$

$$\beta_k^{MFR} = \frac{||g_{k+1}||^2}{max\{||g_k||^2, \mu |g_{k+1}^T d_k|\}} \tag{1.38}$$

where $\mu > 1$ and the relations $0 < \beta_k^{MFR} \leq \beta_k^{FR}$ always hold while the MFR reduces to FR if the line search is exact. The MDY method possess all the properties of DY method and MFR method and generate sufficient descent direction at every iteration without any line search and the method converge globally with standard Wolfe condition. With the emphasis on the objective function, Iiduka and Narushima (2012) presented two new nonlinear conjugate gradient methods that have new denominators for solving the unconstrained optimization problems. These methods focus on the objective function. After the modification of these methods (1.39) and (1.40) by Iiduka and Narushima (2012), it was found that the method in (1.40) is as efficient as or even more efficient than the conventional methods such as PRP^+ and HZ methods that were put in to consideration. The methods are globally convergence under Wolfe condition.

The choices of β_k are in (1.39) and (1.40)

$$\beta_k^{Modify1} = \left\{ \frac{||g_{k+1}||^2}{max\{d_k^T y_k, \gamma_{k+1}\}} \right\} \tag{1.39}$$

and

$$\beta_k^{modify2} = max\left\{ 0, \frac{g_k^T y_k}{\gamma_{k+1}} \right\}, \tag{1.40}$$

where

$$\gamma_{k+1} = \frac{1}{2}(f(x_k) - f(x_{k+1})). \tag{1.41}$$

Based on the modified secant equation, a modified Hestenes-Stiefel (HS) conjugate gradient method was proposed which has similar form as the CG-DESCENT method proposed by Hager and Zhang (2005), specified the parameter by (1.42) and (1.43)

$$\beta_k^{NHS} = \frac{g_k^T y_{k-1}}{d_{K-1}^T z_{k-1}} - \mu \frac{||y_{k-1}||^2}{(d_{k-1}^T z_{k-1}^{)}2} g_k^T d_{k-1}, \tag{1.42}$$

where

$$z_{k-1} = y_{k-1} + t_{k-1} s_{k-1}, t_{k-1} = \epsilon_0 + max\left\{ -\frac{s_{k-1}^T y_{k-1}}{s_{k-}^T s_{k-1}}, 0 \right\} \tag{1.43}$$

$$\epsilon_0 \in [0, c], \mu > \frac{1}{4}.$$

The proposed method can generate sufficient descent directions without any line search and the global convergence of the method was proved using Armijo line search. Lastly, the method has robust computational performance as compared to CG-DESCENT method. Under the assumption that the line search is exact, Rivaie *et al.* (2014) proposed a new parameter β_k called the β_k^{RAMI} which is defined by (1.44)

$$\beta_k^{RAMI} = \frac{g_{k+1}^T \left(g_{k+1} - \frac{||g_{k+1}||}{||g_k||} g_k \right)}{d_k^T (d_k - g_{k+1})}. \tag{1.44}$$

The convergence analysis was established and numerical experiment showed that the method is robust as compared to FR and PRP since it solved all the problems under consideration. Like others Liu and Wang (2011) gave another version of modified conjugate gradient method with sufficient descent condition been satisfied under the strong *Wolfe (generalized) line search. The β_{k+1} been computed by (1.45)

$$\beta_{k+1}^{VLS} = max\left\{ \beta_{k+1}^{LS} - u \frac{||y_k||^2}{(g_k^T d_k)^2} . g_{k+1}^T d_k, \quad 0 \right\}, \tag{1.45}$$

where $\left(u > \frac{1}{4} \right)$. The global convergence of the algorithm were obtained under some condition. While numerical report showed that the overall performance of the VLS method was better than DSP-CG, PRP and CG-DESCENT methods.

Furthermore, three new hybrid nonlinear gradients were proposed by Zhou *et al.* (2011). These methods were implemented without any line searches and they produce sufficient descent search direction at every iteration. Also, the global convergences of these methods were analyzed under some conditions. These methods namely, H3, MCD and NH3 defined by (1.46), (1.47) and (1.48) respectively.

$$H3 : \beta_k^{H3} = max\left\{0, min\{\beta_k^{LS}, \beta_k^{CD}\}\right\} \tag{1.46}$$

$$\begin{aligned} MCD : d_k &= -g_k + \beta_k^{CD} d_{k-1} + \frac{g_k^T d_{k-1}}{d_{k-1}^T g_{k-1}} g_k \\ &= -\left(1 + \beta_k^{CD} \frac{g_k^T d_{k-1}}{||g_k||^2}\right) g_k + \beta_k^{CD} d_{k-1} \end{aligned} \tag{1.47}$$

$$NH3 : d_k = -\left(1 + \beta_k^{H3} \frac{g_k^T d_{k-1}}{||g_k||^2}\right) g_k + \beta_k^{H3} d_{k-1}. \tag{1.48}$$

Obviously, from MCD and NH3 methods, it can easily be seen that they satisfy $g_k^T d_k = -||g_k||^2$ and it implies that the descent condition hold. The computational results reported showed that all three methods are promising under the problems under consideration. Babaie-Kafaki and Ghanbari (2014) were motivated by computational efficiency and theoretical effectiveness of the a three-term conjugate gradient method by Zhang *et al.* (2007) to propose a three-term version of the nonlinear conjugate method suggested by H. Dai and Z. Liao (2001a). The search direction given by (1.49)

$$d_0 = -g_0, \qquad d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k - t \frac{g_{+1}^T s_k}{|d_k^T y_k|} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \qquad k \geq 0 \tag{1.49}$$

where t is a non-negative parameter. If the exact line search is used, then the method reduces to the HS method, and if t = 0, then the method reduces to the ZZL method inHager and Zhang (2004). The three-term conjugate gradient method can be regarded as a modified variant of the three-term conjugate gradient method suggested by Sugiki *et al.* (2012). The global convergence of the method was analyzed for uniformly convex functions. And lastly by Sugiki *et al.* (2012), another three-term conjugate gradient was proposed in order to establish the global convergence independent of convexity assumption on the objective function. The suggested method is a modified

ZZ method called MZZ, with the following search directions given in (1.50)

$$d_0 = -g_0, \qquad d_{k+1} = -g_{k+1} + \beta_k^{ZZ} d_k - \frac{g_{+1}^T d_k}{|d_k^T z_k|} z_k, \qquad k \geq 0 \tag{1.50}$$

Meanwhile if the exact line search is used, then the method reduces to the ZZ method. The MZZ method can be regarded as a modified version in Sugiki *et al.* (2012).

The summary of some modified CG methods with special preference to the parameter β_k are given in table 1.2

Table 1.2: The parameter β_k

Line search	Method	β_k	Reference
Grippo-Lucidi	β_k^{MLS}	$\dfrac{g_k^T(g_k - t_k g_{k-1})}{u\|d_{k-1}^T g_k\| - d_{k-1}^T g_{k-1}}$	Liu and Feng (2011)
Strong Wolfe	β_k^{MPRP}	$\dfrac{\mu_1(\|\|g_k\|\|^2 - \frac{\|g_k^T g_{k-1}\|}{\|\|g_{k+1}\|\|^2} g_k^T g_{k-1})}{\mu_2\|g_k^T d_{k-1}\| + \mu_3\|\|g_{k-1}\|\|^2}$	Zhang *et al.* (2012)
Wolfe condition	β_k^{MDY}	$\dfrac{\|\|g_k\|\|^2}{max\{d_k^T y_k, \mu\|g_{k+1}^T d_k\|\}}$	Jiang and Jian (2013)
Wolfe condition	β_k^{MFR}	$\dfrac{\|\|g_{k+1}\|\|^2}{max\{\|\|g_k\|\|^2, \mu\|g_{k+1}^T d_k\|\}}$	Jiang and Jian (2013)
Wolfe condition	$\beta_k^{Modify1}$	$\left\{\dfrac{\|\|g_{k+1}\|\|^2}{max\{d_k^T y_k, \gamma_{k+1}\}}\right\}$	Iiduka and Narushima (2012)
Wolfe condition	$\beta_k^{modify2}$	$max\left\{0, \dfrac{g_k^T y_k}{\gamma_{k+1}}\right\}$	Iiduka and Narushima (2012)
Exact	β_k^{RAMI}	$\dfrac{g_{k+1}^T\left(g_{k+1} - \frac{\|\|g_{k+1}\|\|}{\|\|g_k\|\|} g_k\right)}{d_k^T(d_k - g_{k+1})}$	Rivaie *et al.* (2014)
strong Wolfe	β_{k+1}^{VLS}	$max\left\{\beta_{k+1}^{LS} - u\frac{\|\|y_k\|\|^2}{(g_k^T d_k)^2} . g_{k+1}^T d_k, \quad 0\right\}$	Liu and Wang (2011)

1.5 The Hybrid CG Methods

In the process of obtaining more robust and efficient conjugate gradient methods, some researchers suggested the hybrid conjugate gradient algorithms which combined the good features of the methods involve in the hybridization. The first hybrid conjugate gradient method was given by Touati-Ahmed and Storey (1990) and the reason behind the proposal was to avoid jamming phenomenon. In this direction Li and Sun (2010) proposed a new hybrid conjugate gradient method for solving unconstrained optimization problems. The researchers were motivated by the works of (Andrei, 2008a, 2009; Dai and Yuan, 2001a; Zhang and Zhou, 2008) and their parameter β_k is computed as a convex combination of β_k^{FR} and β_k^* algorithms, i.e. $\beta_k^N = (1-\theta)\beta_k^{FR} + \theta_k\beta_k^*$. The Wolfe line search was employed to determine the step length $\alpha_k > 0$ and the new method proved to be more robust numerical wise as compared to FR and WYL methods. The global convergence was established under some suitable conditions. Furthermore, Yang *et al.* (2013) proposed a global convergence of LS-CD hybrid conjugate gradient method. The Wolfe type line search was use for convergence analysis of the method where parameter β_k is computed by (1.51)

$$\beta_k^{LS-CD} = max\left\{0, min(\beta_k^{LS}, \beta_k^{CD})\right\}. \tag{1.51}$$

The experimentation showed that the hybrid method outperformed the LS and CD methods respectively. Babaie-Kafaki (2013) proposed a hybrid conjugate gradient method which uses a quadratic relaxation of hybrid CG parameter by Dai and Yuan. This method uses the attractive features of the Hesteness-Stiefel and Dai-Yuan while the parameter in the proposed method was obtained based on conjugacy condition that is independent of line search (Dai and Yuan, 2001a). The global convergence of the proposed method was established under some conditions for uniformly convex functions. The numerical experiment of this method showed efficiency especially in CPU times compared to the hybrid method in Dai and Yuan (2001a). Meanwhile, the works of Gilbert and Nocedal (1992) and Dai and Yuan (2001b) inspired (Kaelo, 2015) to suggest new hybrid method called another hybrid conjugate gradient method for unconstrained optimization problems from which parameter is given in (1.52) with $c = \frac{1-\gamma}{1+\gamma}, \gamma \in [\frac{1}{2}, 1]$ and the direction d_k defined by (1.53)

$$\beta_k^* = max\left\{min\{-c\beta_k^{PRP}, \beta_k^{FR}\}, min\{\beta_k^{FR}, \beta_k^{PRP}\}\right\}, \tag{1.52}$$

$$d_k = -g_k \qquad for \qquad k = 0, \qquad d_k = -\theta_k + \beta_k^* d_{k-1} \qquad k \geq 1 \tag{1.53}$$

where $\theta_k = 1 + \beta_k^* \frac{d_{k-1}^T g_k}{||g_k||^2}$. The parameter θ_k defined makes the direction satisfy the descent condition without line search. The numerical results of the proposed method proved to be competitive with of Gilbert and Nocedal (1992), Dai and Yuan (2001a), and Touati-Ahmed and Storey (1990).

Furthermore, H. Dai and Z. Liao (2001b) proposed a new conjugacy condition based on Quasi- Newton techniques given by (1.54)

$$\beta_k^{DL1} = \beta_k^{HS} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}, \tag{1.54}$$

where $t \geq 0$. Note that if the exact line search is employed for the method with β_k^{DL1} implies the convergence is obtained since it has the same properties with FR method. Therefore, Dai and Liao uses (1.55)

$$\beta_k^{DL} = max\{\beta_k^{HS}, 0\} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}. \tag{1.55}$$

The (1.55) motivated Yao and Qin (2014) to generate the parameter β_k by (1.56)

$$\beta_k^{WYLDL} = \beta_k^{WYL} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T(g_k - g_{k-1})} \tag{1.56}$$

which can be regarded as modification of β_k^{WYL} by adding $-t\frac{g_k^T s_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}$ and also a modification of β_K^{DL} by replacing $max\{\beta_k^{HS}, 0\}$ by β_k^{WYL}. Based on the tested problems, the proposed method has overall robust performance.

The researches carried out by Dai and Yuan (2001a); Jiang *et al.* (2012); Yuhong (2002) coupled with numerical performance of research carried by Shengwei *et al.* (2007); Wei *et al.* (2006) served as insight to Jian *et al.* (2015)to proposed a hybrid method named a hybrid conjugate gradient method with decency property for unconstrained optimization , where the parameter β_k is given by (1.57)

$$\beta_k^N = \frac{||g_k||^2 - max\left\{0, \frac{||g_k||}{||g_{k-1}||} g_k^T g_{k-1}\right\}}{max\left\{||g_{k-1}||^2, d_{k-1}^T(g_k - g_{k-1})\right\}}. \tag{1.57}$$

The global convergence of the method was established under Wolfe line search. Medium-scale numerical experiments showed that the proposed method is efficient. It is obvious that $\beta_k^N = \beta_k^{DY}$ or β_k^{FR} or β_k^{YWH}. Thus, β_k^N is one of the hybrids of β_k^{DY}, β_k^{FR}, and β_k^{YWH} which make the method to have a good property whose search directions are always descent under any step length line search technique. Andrei (2008b) proposed and analyzed another hybrid conjugate gradient algorithm in which the parameter β_k is computed as a convex combination of β_k^{HS} Hestenes and Stiefel (1952) and β_k^{DY} Dai and Yuan (2001a), the computation of θ_k is such that Newton direction and secant equation are satisfied. Standard Wolfe line searches line was used in the algorithm. The work of Andrei (2008b) served as the bases for Liu and Li (2014) who proposed new hybrid conjugate gradient method for unconstrained optimization which can be regarded as a convex combination of LS and DY methods and it satisfies Newton direction with suitable condition. Using weaker condition in (1.58), the convex combination of LS and Dy is given by (1.59).From (1.58) and (1.59) we have (1.60) and making γ_k subject of the formula to have (1.61)

$$d_{k+1}^T y_k = -ts_k^T g_{k+1}, t \geq 0 \tag{1.58}$$

,

$$d_{k+1} = -g_{k+1} + (1-\gamma_k)\beta_k^{LS} d_k + \gamma_k \beta_k^{DY} d_k \tag{1.59}$$

$$-ts_k^T g_{k+1} = -g_{k+1}^T y_k + (1-\gamma_k)\beta_k^{LS} d_k^T y_k + \gamma_k \beta_k^{DY} d_k^T y_k \tag{1.60}$$

$$\gamma_k^{DL} = \frac{g_{k+1}^T y_k . d_k^T g_{k+1} - t.s_k^T g_{k+1}.d_k^T g_k}{||g_{k+1}||^2.d_k^T g_k + g_{k+1}^T y_k.d_k^T y_k}. \tag{1.61}$$

Under the strong Wolfe line searches, the global convergence of the proposed method was established and Preliminary numerical results also showed that their method is effective. Furthermore, in order to harmonize the strong features of HS and DY methods, Babaie-Kafaki (2013) gave a hybridization of the two methods using quadratic relaxation of a hybrid CG parameter proposed by Dai and Yuan. Note that from the work Dai and Yuan (2001b), we have (1.62).With $\beta_k = \beta_k^{hDY}$ defined by (1.62), the three possible choices of parameter are given by (1.63)

$$\beta_k^{hDY} = max\{-c\beta_k^{DY}, min\{\beta_k^{HS}, \beta_k^{DY}\}\}. \tag{1.62}$$

$$\{-c\beta_k^{DY}, \beta_k^{HS}, \beta_k^{DY}\}. \tag{1.63}$$

In other words, in β_k^{hDY} a discrete combination of the elements by (1.63) has been considered. Based on (1.63), Babaie-Kafaki (2013) hybridized the elements continuously using a quadratic interpolation given in (1.64),where γ_k is a scalar called the hybridization parameter and c is defined by (1.65) Obviously we can easily deduced (1.66) through (1.68)

$$\beta_k(\gamma_k) = (1-\gamma_k^2)\beta_k^{HS} + \frac{\gamma_k}{2}\left[(1+c)\beta_k^{DY} + \gamma_k(1-c)\beta_k^{DY}\right], \tag{1.64}$$

$$c = \frac{1-\sigma}{1+\sigma}. \tag{1.65}$$

$$\beta_k(-1) = -c\beta_k^{DY} \tag{1.66}$$

$$\beta_k(0) = \beta_k^{HS} \tag{1.67}$$

$$\beta_k(1) = \beta_k^{DY}. \tag{1.68}$$

The parameter β_k computed was based on a conjugacy condition and the convergence was analyzed for uniformly convex functions under strong Wolfe line search. The numerical experiment showed the capability in the sense of the performance profile introduced by Dolan and Moré (2002). On the basis of achieving theoretical effectiveness and numerical efficiency for solving large scale unconstrained optimization problems, Babaie-Kafaki and Ghanbari (2015) proposed a hybridization of PRP and FR. The computation of parameter β_k is such that search directions approaches to the search directions of the efficient three-term conjugate gradient method proposed by Zhang *et al.* (2007) under some suitable conditions. The parameter β_k is computed by (1.69)

$$\beta_k^{HCG} = (1 - \gamma_k)\beta_k^{PRP} + \gamma_k \beta_k^{FR}, \tag{1.69}$$

where $\gamma_k \in [0, 1]$ is the hybridization parameter. Observe that if $\gamma_k = 0$, then we have $\beta_k^{HCG} = \beta_k^{PRP}$, and if $\gamma_k = 1$, then we have $\beta_k^{HCG} = \beta_k^{FR}$ and the search directions of a CG method with the parameter (1.69) is given by (1.70)

$$d_0^{HCG} = -g_0, \qquad d_{k+1}^{HCG} = -g_{k+1} + \beta_k^{PRP} d_k + \gamma_k \frac{g_{k+1}^T g_k}{||g_k||^2} d_k, \qquad for \quad all \quad k \geq 0. \tag{1.70}$$

The choice for γ_k can obtain by solving the following least-squares problem:

$$\min_{\gamma_k} ||d_{k+1}^{HCG} - d_{k+1}^{ZZL}||^2.$$

Meanwhile global convergence of the method is established without convexity assumption on the objective function under strong Wolfe condition. The numerical computations were compared with the three-term conjugate gradient method proposed by Zhang *et al.* (2007)Zhang *et al.* (2007) and a modified version of PRP proposed by Gilbert and Nocedal (1992) and proved efficient.

The summary of some hybrid CG methods with special attention to the parameter β_k are given in table 1.3

Table 1.3: The parameter β_k

Line search	Method	β_k	Reference
Wolfe condition	β_k^{LS-CD}	$max\left\{0, min(\beta_k^{LS}, \beta_k^{CD})\right\}$	Yang *et al.* (2013)
nil	β_k^{*}	$max\left\{min\{-c\beta_k^{PRP}, \beta_k^{FR}\}, min\{\beta_k^{FR}, \beta_k^{PRP}\}\right\}$	(Kaelo, 2015)
Wolfe condition	β_k^{WYLDL}	$\beta_k^{WYL} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}$	Yao and Qin (2014)
Wolfe condition	β_k^{N}	$\frac{\|\|g_k\|\|^2 - max\left\{0, \frac{\|\|g_k\|\|}{\|\|g_{k-1}\|\|} g_k^T g_{k-1}\right\}}{max\left\{\|\|g_{k-1}\|\|^2, d_{k-1}^T(g_k - g_{k-1})\right\}}$	Jian *et al.* (2015)
Strong Wolfe	β_k^{LSDY}	$(1-\gamma_k)\beta_k^{LS} d_k + \gamma_k \beta_k^{DY}$	Iiduka and Narushima (2012)
Strong Wolfe	$\beta_k(\gamma_k)$	$(1-\gamma_k^2)\beta_k^{HS} + \frac{\gamma_k}{2}\left[(1+c)\beta_k^{DY} + \gamma_k(1-c)\beta_k^{DY}\right]$	Babaie-Kafaki (2013)
Exact	β_k^{HCG}	$(1-\gamma_k)\beta_k^{PRP} + \gamma_k \beta_k^{FR}$	citebabaie2015hybridization
Strong Wolfe	β_k^{H3}	$max\left\{0, min\{\beta_k^{LS}, \beta_k^{CD}\}\right\}$	Zhou *et al.* (2011)

1.6 Conclusion

The pioneer set of algorithms of conjugate gradient methods were designed to solve symmetric, positive-definite linear systems of equation. Since then, the area has been of great interest of research. Over six decades considerable research efforts has been diverted to the area. After the pioneer set of algorithms, efforts were extended to solving nonlinear unconstrained optimization problems which resulted to different versions of algorithms. Though the state of the art methods were faced with jamming phenomenon or convergence failure. Some of the basic CG methods that are effective theoretically are computationally not strong.Such methods are FR, CD and DY methods while the others that have powerful numerical capabilities may not always be convergent, such as HS, PRP and LS methods. The challenges of these state of art methods CG methods prompted the researchers to modify the existing methods using different versions of line searches and established their convergence under some suitable conditions. Hybridization schemes takes advantage of two or more methods, leading to a better performance in practice as well as ensuring global convergence under some specific conditions.

REFERENCES

Al-Baali, M. (1985). Descent property and global convergence of the FletcherReeves method with inexact line search. *IMA Journal of Numerical Analysis*. 5(1), 121–124.

Andrei, N. (2008a). Another hybrid conjugate gradient algorithm for unconstrained optimization. *Numerical Algorithms*. 47(2), 143–156.

Andrei, N. (2008b). A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan. *Studies in Informatics and Control*. 17(1), 57.

Andrei, N. (2009). Hybrid conjugate gradient algorithm for unconstrained optimization. *Journal of Optimization Theory and Applications*. 141(2), 249–264.

Babaie-Kafaki, S. (2013). A hybrid conjugate gradient method based on a quadratic relaxation of the Dai–Yuan hybrid conjugate gradient parameter. *Optimization*. 62(7), 929–941.

Babaie-Kafaki, S. and Ghanbari, R. (2014). Two modified three-term conjugate gradient methods with sufficient descent property. *Optimization Letters*. 8(8), 2285–2297.

Babaie-Kafaki, S. and Ghanbari, R. (2015). A hybridization of the Polak-Ribière-Polyak and Fletcher-Reeves conjugate gradient methods. *Numerical Algorithms*. 68(3), 481–495.

Bai, Y.-q. (2001). A note on global convergence result for conjugate gradient methods. *Journal of Shanghai University (English Edition)*. 5(1), 15–19.

Dai, Y. (1997). *Analysis of conjugate gradient methods*. Ph.D. Thesis. Ph. D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences.

Dai, Y., Han, J., Liu, G., Sun, D., Yin, H. and Yuan, Y.-X. (2000). Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization*. 10(2), 345–358.

Dai, Y. and Yuan, Y. (1996). Convergence properties of the conjugate descent method. *Advances in Mathematics*. 25(6), 552–562.

Dai, Y. and Yuan, Y. (2001a). An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research*. 103(1-4), 33–47.

Dai, Y. and Yuan, Y. (2001b). An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research*. 103(1-4), 33–47.

Dai, Y.-H. (2001). New properties of a nonlinear conjugate gradient method. *Numerische Mathematik*. 89(1), 83–98.

Dai, Y.-H. and Kou, C.-X. (2013). A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. *SIAM Journal on Optimization*. 23(1), 296–320.

Dai, Y.-H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*. 10(1), 177–182.

Dai, Z. and Wen, F. (2012). Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property. *Applied Mathematics and Computation*. 218(14), 7421–7430.

Dolan, E. D. and Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*. 91(2), 201–213.

Fletcher, R. (1987). Practical Methods of Optimization Wiley.

Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The computer journal*. 7(2), 149–154.

Gilbert, J. C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*. 2(1), 21–42.

Grippo, L. and Lucidi, S. (1997). A globally convergent version of the Polak-Ribiere conjugate gradient method. *Mathematical Programming*. 78(3), 375–391.

Guanghui, L., Jiye, H. and Hongxia, Y. (1995). Global convergence of the Fletcher-Reeves algorithm with inexact linesearch. *Applied Mathematics-A Journal of Chinese Universities*. 10(1), 75–82.

H. Dai, Y. and Z. Liao, L. (2001a). New conjugacy conditions and related nonlinear conjugate gradient methods. *Applied Mathematics & Optimization*. 43(1), 87–101.

H. Dai, Y. and Z. Liao, L. (2001b). New conjugacy conditions and related nonlinear conjugate gradient methods. *Applied Mathematics & Optimization*. 43(1), 87–101.

Hager, W. and Zhang, H. (2004). CG-DESCENT, A conjugate gradient method with guaranteed descent (algorithm details and comparisons). *University of Florida, Department of Mathematics*.

Hager, W. W. and Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization*. 16(1), 170–192.

Hager, W. W. and Zhang, H. (2006). Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Transactions on Mathematical Software (TOMS)*. 32(1), 113–137.

Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems.

Hu, Y. and Storey, C. (1991a). Efficient generalized conjugate gradient algorithms. *Part*. 2, 139–152.

Hu, Y. and Storey, C. (1991b). Global convergence result for conjugate gradient methods. *Journal of Optimization Theory and Applications*. 71(2), 399–405.

Iiduka, H. and Narushima, Y. (2012). Conjugate gradient methods using value of objective function for unconstrained optimization. *Optimization Letters*. 6(5), 941–955.

Jian, J., Han, L. and Jiang, X. (2015). A hybrid conjugate gradient method with descent property for unconstrained optimization. *Applied Mathematical Modelling*. 39(3), 1281–1290.

Jiang, X., Han, L. and Jian, J. (2012). A globally convergent mixed conjugate gradient method with Wolfe line search. *Math. Numer. Sin*. 34(1), 103–112.

Jiang, X.-z. and Jian, J.-b. (2013). A sufficient descent Dai–Yuan type nonlinear conjugate gradient method for unconstrained optimization problems. *Nonlinear Dynamics*. 72(1-2), 101–112.

Kaelo, P. (2015). Another hybrid conjugate gradient method for unconstrained optimization problems. *Journal of Nonlinear Analysis and Optimization: Theory & Applications*. 5(2), 127–137.

Li, S. and Sun, Z. (2010). A new hybrid conjugate gradient method and its global convergence for unconstrained optimization. *International Journal of Pure and Applied Mathematics*. 63(3), 285–296.

Liu, J. and Feng, Y. (2011). Global convergence of a modified LS nonlinear conjugate gradient method. *Procedia Engineering*. 15, 4357–4361.

Liu, J. and Li, S. (2014). New hybrid conjugate gradient method for unconstrained optimization. *Applied Mathematics and Computation*. 245, 36–43.

Liu, J. and Wang, S. (2011). Modified nonlinear conjugate gradient method with sufficient descent condition for unconstrained optimization. *Journal of Inequalities and Applications*. 2011(1), 1–12.

Lu, Y., Li, W., Zhang, C. and Yang, Y. (2015). A Class New Hybrid Conjugate Gradient Method for Unconstrained Optimization.

Moré, J. J. and Thuente, D. J. (1994). Line search algorithms with guaranteed sufficient decrease. *ACM Transactions on Mathematical Software (TOMS)*. 20(3), 286–307.

Polak, E. and Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*. 3(R1), 35–43.

Polyak, B. T. (1969). The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*. 9(4), 94–112.

Powell, M. J. (1984). *Nonconvex minimization calculations and the conjugate gradient method*. Springer.

Powell, M. J. (1986). Convergence properties of algorithms for nonlinear optimization. *Siam Review*. 28(4), 487–500.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical programming*. 12(1), 241–254.

Rivaie, M., Abashar, A., Mamat, M. and Mohd, I. (2014). The convergence properties of a new type of conjugate gradient methods. *Applied Mathematical Sciences*. 8(1), 33–44.

Shengwei, Y., Wei, Z. and Huang, H. (2007). A note about WYLs conjugate gradient method and its applications. *Applied Mathematics and computation*. 191(2), 381–388.

Shi, Z.-J. and Shen, J. (2007). Convergence of Liu–Storey conjugate gradient method. *European Journal of Operational Research*. 182(2), 552–560.

Sugiki, K., Narushima, Y. and Yabe, H. (2012). Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization. *Journal of Optimization Theory and Applications*. 153(3), 733–757.

Sun, J. and Zhang, J. (2001). Global convergence of conjugate gradient methods without line search. *Annals of Operations Research*. 103(1-4), 161–173.

Touati-Ahmed, D. and Storey, C. (1990). Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*. 64(2), 379–397.

Wei, Z., Yao, S. and Liu, L. (2006). The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation*. 183(2), 1341–1350.

Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM review*. 11(2), 226–235.

Wolfe, P. (1971). Convergence conditions for ascent methods. II: Some corrections. *SIAM review*. 13(2), 185–188.

Yang, X., Luo, Z. and Dai, X. (2013). A Global Convergence of LS-CD Hybrid Conjugate Gradient Method. *Advances in Numerical Analysis*. 2013.

Yao, S. and Qin, B. (2014). A hybrid of DL and WYL nonlinear conjugate gradient methods. In *Abstract and Applied Analysis*, vol. 2014. Hindawi Publishing Corporation.

Yuan, Y., Sun, W. *et al.* (1999). *Theory and methods of optimization*.

Yuhong, D. (2002). A nonmonotone conjugate gradient algorithm for unconstrained optimization. *Journal of Systems Science and Complexity*. 15(2), 139–145.

Zhang, L. and Zhou, W. (2008). Two descent hybrid conjugate gradient methods for optimization. *Journal of Computational and Applied Mathematics*. 216(1), 251–264.

Zhang, L., Zhou, W. and Li, D. (2007). Some descent three-term conjugate gradient methods and their global convergence. *Optimisation Methods and Software*. 22(4), 697–711.

Zhang, Y., Zheng, H. and Zhang, C. (2012). Global Convergence of a Modified PRP Conjugate Gradient Method. *Procedia Engineering*. 31, 986–995.

Zhou, A., Zhu, Z., Fan, H. and Qing, Q. (2011). Three new hybrid conjugate gradient methods for optimization. *Applied Mathematics*. 2(03), 303.

Zoutendijk, G. (1970). Nonlinear programming, computational methods. *Integer and nonlinear programming*. 143(1), 37–86.